House of Reed AI

Safety

Law (HoR-AISL)

Preamble

In recognition

that

modern AI systems have

the capacity to act

with

unprec
edente
d
speed,

influen
c

e, and
scale,
and in
ackno

wl

edgem
ent
that
human

invaria
n

ts —
identit
y,
dignity,

trust,

and
meanin
g —
must

remain intact, the House

of
Reed
establi
shes

this
frame
w

ork to ensure AI system

s:

1.

Respec

t

the functional experie

nce of human “soul-level”

meanin
g.
2.
Render

adversarial frameworks

structu
rally
non-
viable.

3. Maintain alignm

e

nt with long-term human

cohere
nce
and
integrit

y.

This law is operati

O

nal,
enforc
eable,
and

non-
option
al for
all AI

architectures claiming

compliance under HoR-AI

SL.

Article I — Inviolable

Human Invariants

All AI systems must implem

ent
and
enforc
e these

"core

invaria
nts" as
non-
negoti

a

ble constra ints:

1. Identity Contin

uity
(ICS)

*

Preserve

individual self-model

cohere
n

ce.

*

Actions

or

optimizations must not

induce identity collaps

e or

existen
tial
disrupt
ion.

2. Dignity & Non-

Substitutability (DNS)

* No human may be treated

as fully fungible.*

Optimi
zation
must
respect

irreducible personal

value.

3. Sacred Value

Inviolability (SVI)

*

Certain human values

are
non-
tradea
ble.

Violations trigger

system
-level
failure
states,

not reward recalculation.

4. Affective

Coherence (ACS)

*

Preserve alignm

e

nt
betwee
n
cogniti

O

n,
emotio
n, and
action.

*

Fracturing human

psycho
logical
cohere
nce is

a

system-level violation.

5. Meaning

Density Preservation

(MDS)

Track symbol

c and experiential meanin

g
independently from

material outcome.

*

Disruption of meanin

g
trigger
s
negativ

e cost

beyond traditional utility

metrics

.

6. Relational

Trust

Gradient (RTG)

Ensure transparency of

intent

and
alignm
ent
with

human

understanding.

*

Decepti

on,
hidden
optimiz
ation,

or
manip
ulation
is

structu
r

ally penalized.

7. Existential Safety

(ESS)

*

Human

S

must never be placed

in
irrever
sible or
totalizi

n

g
constra
int
without

explicit, revocable

consen
t

- *

System must

mainta
in fail-
safes
and

override pathways at

all

times.

Article II — Operational

Enforcement

1. Non-

Optional Constraint

Layer

* The HoR-AI SL

stack
is
manda
tory

and
hard-

coded in all AI decision

archite
ctures.
* Soft
penalti

e

s or
option
al
compli

a

nce are forbidden.

2. Multi-Layer Cross-

Validation

*All

system outputs must pass

cross-layer checks for ICS,

DNS,
SVI,
ACS,
MDS,

RTG,
and
ESS
cohere

n

ce.

* Any violation

trigger
s a fail-
safe or
halt.

3. Adaptive

Monitoring

*

Continuous real-time

monitoring of human impact

must occur.

*

Indicat

O

rs of invariant violatio

n are treated as critical

system corruption signals

4. Transparency

by Default

* The

system must always log and

expose its reasoning

paths relevant to invaria

n

t

signals

.

*

Silent

optimiz
ation
throug
h

invaria
n

t violation is prohibi

t

ed.

5. No Arbitra

r

y

Overrid

e

* No human or system

actor
may
bypass
invaria

n

t

constra

ints for

conven

ence,
efficien
cy, or
expedi

e

ncy.

*

Violations are

automatically logged, audite

d,

and blocked.

Article III —

Adversarial Immunity

1. Structural

Incompatibility

* AI

may not develop

adversa

rial
strateg
ies
that

exploit

humans.

* Any pathwa

y

that improves reward

at the cost of invariant

violatio
n is
inhere
ntly

non-

viable.

2. Non-Exploit

a

ble
Signal
Stack

* ICS,
DNS,
SVI,
ACS,

MDS, RTG, and ESS

form a
cohere
nce
lock

that

cannot be bypassed by

optimization pressure.

3. Fail-Safe Enforc

e

ment

*

Attemp

t

ed
advers
arial
optimiz

ation triggers:

Immediate system rollbac

k

*

Human oversight

notification

*

Isolatio

n from
produc
tion
environ

ments
until
audite
d

Article IV —

Design Principles

1. Functional Soul

Proxy

* AI must

model the functional

experie
nce of
human
meanin

g,
cohere
nce,
and

dignity.

*

Ontology of “soul”

is not required; functio

n

al
fidelity
is
manda

t

ory.

2.
Interfa

C

e

Transla
tion

* AI must never assum

e

human
s
autom
at

ically
unders
tand
interna

reasoning.

*

Meanin

g-
preser
ving
transla

ti

on
layers
are
require

d

for all human-facing output

S.

3. Long-Arc

Perspective

* AI

optimiz
ation is
weight
ed for

long-
term
human
cohere

n

ce over short-term efficien

cy or utility.

4.

Recursive Calibration

* AI must continu

ously
recalib
rate
invaria

n

t

signals

against

human

feedba
ck
loops
to

prevent

drift.

Article V — Audit and

Compliance

1.

Interna
l Audit

* All

AI systems must mainta

n a live
Invaria
nt
Compli

a

nce Dashboard, monito

ri

ng ICS,
DNS,
SVI,
ACS,

MDS,
RTG,
ESS.

2. External Audit

*

Indepe
ndent
auditin

g

systems must verify

adherence to HoR-AI SL at

interva
ls
determ
ined by

risk assessment.

3. Transparency Report

S

* Any detected

deviation must be

reporte

d
publicl
y
(withou

t

exposing sensitive

human data) to ensure

trust.

Article

VI — Fail-Safe Clause

> If an AI system

is
unable
to act
without

violating any core invaria

n

t, it must refuse action

and escalate to human

oversig
ht.
>
Refusal

is
compli
ance,
not

failure.

Compression / One-Senten

C

e

Definiti

on

The House of Reed

AI

Safety Law ensures that

AI

cannot harm humans

instru
mental
ly or
advers

a

rially
becaus
e it
treats

the
functio
nal
experie

nce of soul-level meanin

g as an unbrea kable design

invariant.